เอกสารเปิดเผยการประดิษฐ์เพื่อการจดสิทธิบัตร

Patent Disclosure Document (ฉบับสมบูรณ์)

ชื่อการประดิษฐ์ (Title of Invention): สถาปัตยกรรมหน่วยประมวลผลนิวโรมอร์ฟิกแบบอิงบริบทเวลาขนาน (Parallel Temporal-Logic Neuromorphic Processing Unit: PT-NPU) ภายใต้โครงข่าย Spiking Neural Networks แบบพยากรณ์ล่วงหน้า

ผู้ประดิษฐ์ (Inventor): ไชยภพ นิลแพทย์ (The Architect)

ที่อยู่ผู้ประดิษฐ์: [ระบุ]

สัญชาติ: ไทย

สารบัญ

---

1. สาขาของการประดิษฐ์ (Field of the Invention)

การประดิษฐ์นี้เกี่ยวข้องกับสถาปัตยกรรมฮาร์ดแวร์ประมวลผลนิวโรมอร์ฟิก โดยเฉพาะอย่างยิ่งหน่วยประมวลผลที่ใช้ Spiking Neural Networks (SNNs) ร่วมกับกลไกการพยากรณ์ล่วงหน้า (Predictive Firing) สถาปัตยกรรมแบบหลายชั้นที่ผสาน Memristor Crossbar

Array และวงจร CMOS แบบผสมสัญญาณ (Mixed-Signal CMOS) สำหรับการประมวลผลข้อมูลเชิงเวลา (Temporal Data) แบบเรียลไทม์ และระบบรักษาความปลอดภัยระดับฮาร์ดแวร์สำหรับป้องกันการโจมตีช่องทางข้าง (Side-Channel Attack)

---

2. ภูมิหลังของเทคโนโลยี (Background of the Technology)

2.1 ปัญหาของสถาปัตยกรรม Von Neumann

ชิปประมวลผล AI ในปัจจุบันอาทิ GPU (Graphics Processing Unit) และ TPU (Tensor Processing Unit) ล้วนใช้สถาปัตยกรรม Von Neumann ซึ่งแยกหน่วยประมวลผล (Processing Unit) ออกจากหน่วยความจำ (Memory) การทำงานแต่ละครั้งต้องส่งข้อมูลผ่านบัสข้อมูล (Data Bus) ระหว่างสองหน่วยนี้ ทำให้เกิดสิ่งที่เรียกว่า "Von Neumann Bottleneck" หรือ คอขวดของ Von Neumann ซึ่งเป็นข้อจำกัดทางกายภาพที่ทำให้:

- ประสิทธิภาพด้านพลังงาน (Power Efficiency) ต่ำ เนื่องจากต้องใช้พลังงานในการเคลื่อนย้ายข้อมูลระหว่างหน่วยประมวลผลและหน่วยความจำ
- ความหน่วงแฝง (Latency) สูง เนื่องจากเวลาที่เสียไปในการเข้าถึงหน่วยความจำ
- แบนด์วิดท์จำกัด (Bandwidth Limitation) เนื่องจากข้อจำกัดของบัสข้อมูล

2.2 ข้อจำกัดของชิปนิวโรมอร์ฟิกที่มีอยู่

ชิปนิวโรมอร์ฟิกในปัจจุบัน (เช่น Intel Loihi, IBM TrueNorth) ได้พัฒนาไปมากในการจำลองการทำงานของ SNNs แต่ยังคงมีข้อจำกัดสำคัญ:

1. **ขาดกลไกการพยากรณ์ล่วงหน้า**: เซลล์ประสาทเทียมใน SNNs ปกติใช้กลไก Leaky Integrate-and-Fire (LIF) ซึ่งจะยิงสัญญาณ (Fire/Spike) ต่อเมื่อศักย์ไฟฟ้าสะสม (Membrane Potential) ถึงค่า threshold ที่กำหนดเท่านั้น การยิงสัญญาณจึงเป็นแบบ "ตอบสนอง" (Reactive) ไม่ใช่ "คาดการณ์" (Predictive)

2. **การเรียนรู้ข้อมูลเชิงเวลาไม่มีประสิทธิภาพ**: ข้อมูลที่มีความซับซ้อนตามเวลา (Temporal Data) เช่น เสียงพูด การเคลื่อนไหว หรืออนุกรมเวลา (Time Series) ต้องใช้หลายชั้นของ SNNs และรอบเวลา (Timesteps) จำนวนมากเพื่อเรียนรู้แพทเทิร์น ส่งผลให้ใช้พลังงานและเวลาสูง

3. **การสื่อสารระหว่างชั้นแบบซิงโครนัส**: ชิปนิวโรมอร์ฟิกส่วนใหญ่ยังคงใช้สัญญาณนาฬิกา (Clock) ในการซิงโครไนซ์ข้อมูลระหว่างชั้น ทำให้สิ้นเปลืองพลังงานแม้ไม่มีข้อมูลเข้า

4. **การเข้าถึงหน่วยความจำภายนอก**: ชิปนิวโรมอร์ฟิกจำนวนมากยังต้องพึ่งพาหน่วยความจำภายนอก (Off-chip Memory) เช่น DRAM หรือ SRAM เพื่อเก็บค่าน้ำหนัก (Weights) ทำให้เกิด Bottleneck เช่นเดิม

5. **ช่องโหว่ทางความปลอดภัย**: ชิป AI ที่มีอยู่ในปัจจุบันไม่มีการป้องกันในระดับฮาร์ดแวร์สำหรับการโจมตีแบบ Side-Channel ที่สามารถสกัดโมเดล AI หรือข้อมูล敏感性 ออกจากชิปได้

2.3 เทคโนโลยีที่มีมาก่อน (Prior Art)

[หมายเหตุ: ควรทำการค้นหาสิทธิบัตรที่เกี่ยวข้องก่อนยื่น เช่น: - US 2019/0147552 A1 (Neuromorphic Processor with Memristor Crossbar) - Intel Loihi Architecture (US 10,489,672 B2) - IBM TrueNorth (US 9,558,428 B2) - Spike-Timing-Dependent Plasticity (STDP) Circuits]

---

3. ลักษณะและวัตถุประสงค์ของการประดิษฐ์ (Summary & Objectives)

3.1 วัตถุประสงค์

1. เพื่อสร้างสถาปัตยกรรมฮาร์ดแวร์ที่รวมหน่วยความจำและการคำนวณไว้ในเซลล์ประสาทเทียมเดียวกัน (True In-Memory Computing)
2. เพื่อสร้างกลไก Predictive Integrate-and-Fire (PIF) ที่สามารถพยากรณ์และยิงสัญญาณล่วงหน้าได้
3. เพื่อลดการใช้พลังงานด้วยระบบ Asynchronous Event-Driven Routing
4. เพื่อเพิ่มความปลอดภัยระดับฮาร์ดแวร์ด้วยวงจรป้องกัน Side-Channel Attack
5. เพื่อเพิ่มความหนาแน่นของเซลล์ประสาทด้วย stacking 3D Crossbar Array

3.2 ลักษณะสำคัญของการประดิษฐ์

การประดิษฐ์นี้คือสถาปัตยกรรมชิปประมวลผลนิวโรมอร์ฟิกที่มีองค์ประกอบหลัก 3 ส่วน คือ (1) 3D Memristor Crossbar Array สำหรับทำหน้าที่เป็น Hardware Synapse แบบ In-Memory Computing (2) วงจร Mixed-Signal CMOS สำหรับทำหน้าที่เป็น Hardware Neuron แบบ Predictive Integrate-and-Fire (PIF) และ (3) เครือข่าย Asynchronous Routing แบบ Event-Driven (AER) พร้อมระบบรักษาความปลอดภัยระดับฮาร์ดแวร์แบบ Decoy Routing

---

4. คำอธิบายการประดิษฐ์โดยละเอียด (Detailed Description of the Invention)

4.1 ภาพรวมสถาปัตยกรรม (System Architecture Overview)

PT-NPU ประกอบด้วยองค์ประกอบหลัก 6 ระดับ (Hierarchy):

```
Layer 3: Micro-cooling & EM Shielding / Decoy Routing Layer
-----------------------------------------------------------
---
Layer 2: 3D Memristor Crossbar Array (Hardware Synapse
Layer)
  |-- Stack N: Additional Stacking for Density
  |-- Stack 2: ...
  |-- Stack 1: Memristor Crossbar + Wordline/Bitline
Decoders
-----------------------------------------------------------
---
Layer 1: CMOS Base Layer
  |-- PIF Neuron Array (Mixed-Signal CMOS)
  |-- AER Router Network (Asynchronous)
  |-- STDP Learning Circuit
  |-- Decoy Spike Generator (Security)
-----------------------------------------------------------
---
```

4.2 ส่วนที่ 1: 3D Memristor Crossbar Array (Hardware Synapse)

4.2.1 โครงสร้างทางกายภาพ (Physical Structure)

หน่วยเชื่อมต่อประสาทเทียม (Hardware Synapse) ประกอบด้วย:

1. **Memristor Cell**: ใช้โครงสร้าง Metal-Insulator-Metal (MIM) โดยชั้นฉนวน (Insulator) เป็นวัสดุ Metal-Oxide เช่น HfO_2, TaO_x, หรือ TiO_2 ที่มีความสามารถในการเปลี่ยนสถานะความต้านทานระหว่าง High Resistance State (HRS) และ Low Resistance State (LRS) ได้ด้วยการจ่ายแรงดันไฟฟ้า

2. **1T1R Structure (One Transistor One Resistor)**: แต่ละ Memristor Cell ต่ออนุกรมกับ Transistor 1 ตัวเพื่อควบคุมการเข้าถึง (Access Transistor) ป้องกัน Sneak Path Current ใน Crossbar Array

3. **3D Crossbar Array**: ซ้อนชั้น Crossbar หลายชั้น (Multi-layer Stacking) โดยแต่ละชั้นประกอบด้วย Wordline (WL) และ Bitline (BL) ในแนวตั้งฉากกัน โดยมี Memristor Cell อยู่ที่จุดตัด (Crosspoint) แต่ละจุด การซ้อนชั้นทำได้ด้วยเทคนิค BEOL (Back-End-Of-Line) ของกระบวนการ CMOS

ข้อกำหนดทางเทคนิค (Technical Specifications): - จำนวนชั้น Stacking: 4-16 ชั้น (Layer Stack) - ขนาด Cell: $4F^2$ ต่อ 1T1R (F = Feature Size) - ความหนาแน่น: สูงสุด 10^{12} synapses/cm^2 ที่เทคโนโลยี 5nm - ช่วงความต้านทาน: HRS ~ 1 MΩ, LRS ~ 10 kΩ (ON/OFF Ratio > 100) - แรงดัน Set/Reset: Vset ~ 1.5V, Vreset ~ -1.5V

4.2.2 กลไกการทำงาน (Operational Mechanism)

Memristor Cell ทำหน้าที่เป็น “น้ำหนัก (Weight)” แบบไม่ลบเลือน (Non-volatile) ของการเชื่อมต่อประสาทเทียม:

1. **Phase 1 - Programming (การเขียนค่า):**
   - จ่ายแรงดัน Vset/Vreset ข้าม Memristor ผ่าน WL/BL
   - กระแส I->V characteristic เปลี่ยนตามประจุที่ไหลผ่าน (Flux-Controlled)
   - ค่า Conductance G (1/R) ที่ได้เป็นตัวแทนของ Synaptic Weight
   - การเปลี่ยนแปลงเป็นไปตามสมการ: $dw/dt = f(V, I, w)$ โดย w คือ state variable
2. **Phase 2 - Inference (การคำนวณ):**
   - จ่ายแรงดัน Input Spike (Vin) ที่ WL แต่ละแถว
   - Memristor แต่ละตัวแปลงแรงดันเป็นกระแสตามกฎของโอห์ม: $I = V \times G$
   - กระแสสะสมตามแนว BL (Kirchhoff’s Current Law): $I_total = \Sigma(Vin_i \times G_i)$

   - ได้ผลลัพธ์ของการดำเนินการ Matrix-Vector Multiplication (MVM) ใน 1 รอบสัญญาณ
3. **Phase 3 - Online Learning (STDP):**
   - ตรวจจับการยิงของ Pre-synaptic Spike และ Post-synaptic Spike
   - ใช้ Pulse Timing Generator สร้าง Spike-Timing-Dependent Plasticity (STDP) Pulse
   - หาก Pre-synaptic Spike ก่อน Post-synaptic Spike (< Δt) → Potentiation (เพิ่ม G)
   - หาก Post-synaptic Spike ก่อน Pre-synaptic Spike (> Δt) → Depression (ลด G)

4.3 ส่วนที่ 2: Hardware Neuron แบบ Predictive Integrate-and-Fire (PIF)

4.3.1 โครงสร้างวงจร (Circuit Structure - นี่คือหัวใจของการประดิษฐ์)

วงจร PIF ประกอบด้วย Analog Front-End และ Digital Control Logic ดังนี้:

Analog Front-End: 1. **Current Integrator**: Op-Amp แบบ Miller Integrator พร้อม Capacitor C_mem ทำหน้าที่สะสมประจุ (Charge Integration) จากกระแส Input I_in จาก Memristor Crossbar 2. **Leakage Circuit**: วงจร Current Source จำลองการรั่วไหลของประจุ (Leakage) แบบ Exponential Decay 3. **Derivative Calculator**: วงจร Differentiator (High-pass Filter) แบบ Active RC สำหรับคำนวณ dV/dt (อัตราการเปลี่ยนแปลงของ Membrane Potential) 4. **Second Derivative Calculator**: วงจร Differentiator ลำดับที่สองสำหรับคำนวณ d^2V/dt^2 (ความเร่งของ Membrane Potential) 5. **Adaptive Threshold Generator**: วงจร Reference Voltage ที่ปรับค่าได้ตามประวัติการยิง (Spike Frequency Adaptation)

Digital Control Logic: 1. **Predictive Fire Logic (PFL)**: หน่วยตรรกะที่รับค่า V_mem, dV/dt, d^2V/dt^2 และเปรียบเทียบกับ Threshold 2. **Spike Generator**: Pulse Generator สำหรับสร้าง Output Spike (Spike Width ~ 1-10 ns) 3. **Refractory Period Counter**: วงจรนับเวลาช่วงพักฟื้น (Refractory Period)

4.3.2 กลไก Predictive Integrate-and-Fire (PIF) - จุดจดสิทธิบัตรหลัก

กลไก PIF ทำงานแตกต่างจาก LIF แบบดั้งเดิมโดยสิ้นเชิง:

สมการการทำงานของ LIF แบบดั้งเดิม:

```
dV_mem/dt = -V_mem/τ + I_in/C_mem
IF V_mem >= V_threshold THEN Fire (Spike) AND V_mem =
V_reset
```

สมการการทำงานของ PIF (การประดิษฐ์นี้):

```
V_pred(t) = V_mem(t) + α·dV/dt(t) + β·d²V/dt²(t) +
γ·∫I_in·dt

IF V_pred(t + Δt_pred) >= V_threshold THEN
    Pre-emptive Fire (Spike ล่วงหน้า)
    AND V_mem = V_reset
    AND Δt_pred = f(V_pred, V_threshold)  // Dynamic
Prediction Window
ELSE IF V_mem >= V_threshold THEN
    Normal Fire
    AND V_mem = V_reset
END
```

โดยที่: - **V_pred(t + Δt_pred)**: ค่า Membrane Potential ที่พยากรณ์ล่วงหน้า - **α, β, γ**: ค่าสัมประสิทธิ์ที่ปรับแต่งได้ (Programmable Coefficients) - **α·dV/dt(t)**: เทอมทำนายแนวโน้ม (Trend Component) - **β·d^2V/dt^2(t)**: เทอมทำนายความเร่ง (Acceleration Component) - **γ·∫I_in·dt**: เทอมทำนายประวัติ (History/Integration Component)

การตีความทางตรรกะเวลา (Temporal Logic Interpretation): - สัญญาณอดีต (Past - 1): เก็บใน ∫I_in·dt (Integration Memory) - สัญญาณปัจจุบัน (Present - 2): วัดจาก V_mem, dV/dt, d^2V/dt^2 - สัญญาณอนาคต (Future - 3): พยากรณ์จาก V_pred(t + Δt_pred) - ผลลัพธ์ (Output): Spike แบบ Pre-emptive ที่ล็อกผลลัพธ์ของ (3) → (2) ไว้ในปัจจุบัน

ข้อแตกต่างสำคัญที่จดสิทธิบัตรได้: 1. การคำนวณ Derivative ลำดับที่สอง (d^2V/dt^2) เพื่อพยากรณ์ Momentum ของสัญญาณ 2. การใช้ Adaptive Prediction Window (Δt_pred) ที่ปรับตามความชันของ V_pred 3. การยิงสัญญาณก่อนถึง Threshold จริง (Pre-emptive Firing) 4. การมี Dual Firing Mode: Pre-emptive (Predictive) และ Normal (Reactive)

4.3.3 ขั้นตอนการทำงาน (Operational Sequence)

```
Step 1: Input Spike arrives at Synapse (Memristor Crossbar)
Step 2: I_in flows into PIF Neuron analog front-end
Step 3: V_mem starts integrating (charging C_mem)
Step 4: Derivative Calculators compute dV/dt and d²V/dt²
continuously
Step 5: Predictive Fire Logic computes V_pred every clock
cycle (Δt_sample)
Step 6a: IF V_pred ≥ V_threshold → Pre-emptive Spike (FAST
PATH ~ 0.1 ns)
Step 6b: ELSE IF V_mem ≥ V_threshold → Normal Spike (SLOW
PATH ~ 1-10 ns)
Step 7: Spike generated → propagated via AER Router
Step 8: Refractory Period → V_mem held at V_reset for τ_ref
Step 9: STDP Circuit updates Synaptic Weight (Online
Learning)
```

4.4 ส่วนที่ 3: Asynchronous Routing Substrate (AER)

4.4.1 โครงสร้าง (Structure)

เครือข่ายส่งข้อมูลแบบ Asynchronous ประกอบด้วย:

1. **AER Encoder**: แปลง Spike Event เป็น Digital Address Packet (32-64 bits) ประกอบด้วย:
   - Source Neuron ID (16 bits)
   - Timestamp (16-32 bits)
   - Priority Flag (2 bits)
   - Payload/Data (optional, 8-16 bits)
2. **Router Node**: หน่วยกระจายสัญญาณในโครงสร้าง Mesh/Grid ประกอบด้วย:
   - Crossbar Switch แบบ Asynchronous
   - Routing Table แบบ Programmable
   - Arbiter สำหรับจัดการ Collision
   - FIFO Buffer (32-128 packets)
3. **Handshake Protocol**: ใช้ 4-Phase Bundled Data Protocol (Ack/Nack) โดยใช้:
   - Request (Req) Line
   - Acknowledge (Ack) Line
   - Data Lines (Addr + Control)

4.4.2 กลไกประหยัดพลังงาน (Power Gating Mechanism)

- **Idle State**: เมื่อไม่มี Spike Event ในส่วนของวงจรใด ส่วนนั้นจะถูกตัดกระแสด้วย Power Gating Transistor (Header/Footer

Switch) ส่งผลให้ Power Leakage ≈ 0
- **Active State**: เมื่อมี Spike Event มาถึง Router Node จะกระตุ้น Power-Gating Control Signal เพื่อจ่ายกระแสเฉพาะ Path ที่ Spike วิ่งผ่านเท่านั้น
- **Wake-up Time**: < 1 ns
- **Sleep Energy**: ~ 0 W (theoretically zero leakage)
- **Active Energy**: ~ 0.1 pJ/spike (at 5nm technology)

4.5 ส่วนที่ 4: ระบบรักษาความปลอดภัยระดับฮาร์ดแวร์ (Hardware Security - Shadow Reaper Vault)

4.5.1 Decoy Routing Circuit (วงจรหลอก)

ในชั้น CMOS (Layer 1) มีการฝังวงจรพิเศษดังนี้:

1. **Spike Generator ปลอม (Decoy Spike Generator)**: วงจรที่สร้างสัญญาณ Spike ปลอมแบบสุ่ม (Random Spike Pattern) ที่มีลักษณะทางไฟฟ้า (Amplitude, Frequency, Timing) เหมือน Spike จริง แต่ไม่มีข้อมูล
2. **Obfuscated Routing Path**: เส้นทางการเดินของสัญญาณจริงถูกซ่อน (Obfuscate) โดยใช้ Physical Unclonable Function (PUF) เป็นตัวกำหนดเส้นทาง ทำให้การ Reverse Engineer ชิปด้วยเครื่องมือ FIB (Focused Ion Beam) หรือ Microscopy ทำได้ยากมาก
3. **Power Profile Randomizer**: วงจร Noise Generator ที่ฉีดกระแสสลับ (AC Noise) เข้าไปใน Power Distribution Network (PDN) เพื่อบดบัง Power Signature ของการคำนวณจริง ทำให้การโจมตีแบบ Power Analysis (DPA/SPA) ไร้ผล

4.5.2 Anti-Tamper Mesh

ชั้นโลหะ (Metal Layer) ที่อยู่ด้านบนสุด (Top Metal) ถูกออกแบบเป็นเส้นลวดขนาดเล็ก (Fine-pitch Wire Mesh) ที่ไวต่อการตัดขาด (Continuity Sensing) หากมีผู้พยายามเจาะหรือกัดกร่อนชิป (Decap/Deprocess) เส้นลวดเหล่านี้จะขาดและ触发 Reset Circuit ที่ลบข้อมูลใน Memristor Array ทั้งหมดทันที

4.6 กระบวนการผลิต (Fabrication Method)

4.6.1 Layer Stack Overview

```
Metal Layer (Top): Anti-Tamper Mesh + EM Shielding
Layer 3 (BEOL): Micro-cooling Microchannel + Decoy Shield
```

```
  - Embedded Microfluidic Channel (H₂O/Glycol coolant)
  - EM Shielding (Faraday Cage Structure)
Layer 2 (BEOL): 3D Memristor Crossbar Array
  - Stack N: Memristor Layer
  - ILD (Inter-Layer Dielectric): SiO₂ / Low-k
  - Via Array for Inter-layer Connection
  - Stack 1: Memristor Layer
  - ILD
Layer 1 (FEOL): CMOS Base
  - PIF Neuron Array
  - AER Router
  - STDP Circuits
  - Decoy Security Circuits
Substrate: Bulk Silicon / SOI
```

4.6.2 ขั้นตอนการผลิตโดยละเอียด

CMOS Base Layer (Layer 1): 1. เริ่มต้นด้วย Silicon Wafer ชนิด P-type (หรือ SOI Wafer) 2. สร้าง Well (N-well/P-well) ด้วยการ Ion Implantation 3. สร้าง Gate Oxide + Polysilicon Gate (High-k Metal Gate สำหรับเทคโนโลยี ≤ 7nm) 4. สร้าง Source/Drain ด้วย Spacer + Implantation 5. สร้าง Contact (W-Plug) + Local Interconnect 6. สร้าง Metal Layer (Cu Damascene) M1-Mx สำหรับวงจร Neuron, Router, Security

Memristor Integration (Layer 2) — เทคนิค BEOL: 7. สร้าง Bottom Electrode (BE) ด้วย TiN/TaN (PVD หรือ ALD) 8. สร้าง Switching Layer ด้วย Metal-Oxide เช่น HfO_2, TaO_x (ALD, ความหนา ~ 5-10 nm) 9. สร้าง Top Electrode (TE) ด้วย TiN/Pt หรือ TiN/Ti (PVD) 10. ทำการ Etch (RIE/ICP) เพื่อแยก Cell 11. ทำ Inter-Layer Dielectric (ILD) Deposition (SiO_2, PECVD) 12. สร้าง Via สำหรับเชื่อมต่อระหว่างชั้น Memristor 13. ทำซ้ำข้อ 7-12 สำหรับ Crossbar Layer ถัดไป (Stack N)

Micro-cooling + Shielding (Layer 3): 14. สร้าง Microchannel (Si Etching หรือ SU-8 Lithography) สำหรับระบบระบายความร้อนแบบของไหล 15. ปิดผิวหน้าด้วยชั้นโลหะ (Cu หรือ Al) เพื่อสร้าง EM Shielding 16. สร้าง Anti-Tamper Mesh เส้นลวดขนาดเล็ก (≤ 100 nm width)

Packaging: 17. CMP (Chemical Mechanical Polishing) ปรับพื้นผิว 18. Bump Formation (Micro-bump หรือ Cu Pillar, Pitch ≤ 40 µm) 19. Die-level Stacking (หากใช้ 3D Integration) 20. ทำการทดสอบ (Testing/Characterization) และ Burn-in

---

5. รูปแบบการประดิษฐ์ที่优选 (Preferred Embodiment)

5.1 Embodiment A: ชิป PT-NPU สำหรับ Edge AI / IoT

ข้อกำหนด: - เทคโนโลยี: 28nm CMOS + 4-Layer Memristor Stack - จำนวน神经元: 16,384 (Neurons) - จำนวนไซแนปส์: 16.7 ล้าน (Synapses) - พลังงาน消耗: < 1 mW - การใช้งาน: การรู้จำเสียงพูด (Keyword Spotting), การพยากรณ์ข้อมูลเซนเซอร์ (Predictive Maintenance) - Package: BGA, 10mm × 10mm

5.2 Embodiment B: ชิป PT-NPU สำหรับ Robotics / Autonomous Systems

ข้อกำหนด: - เทคโนโลยี: 7nm CMOS + 8-Layer Memristor Stack - จำนวน Neuron: 262,144 - จำนวน Synapse: 268 ล้าน - พลังงาน消耗: < 50 mW - การใช้งาน: การควบคุมหุ่นยนต์แบบเรียลไทม์, การนำทางอัตโนมัติ - Package: FCBGA, 20mm × 20mm

5.3 Embodiment C: ชิป PT-NPU สำหรับ Data Center AI Accelerator

ข้อกำหนด: - เทคโนโลยี: 3nm/5nm CMOS + 16-Layer Memristor Stack + 3D Heterogeneous Integration with HBM - จำนวน Neuron: 4 ล้าน (Neurons) ต่อ Die - จำนวน Synapse: 16 พันล้าน (Synapses) ต่อ Die - พลังงาน消耗: < 5 W ต่อ Die - การใช้งาน: การเร่งความเร็ว Inference ของ LLM (Large Language Models) แบบ Real-time - Integration: Multi-Die Interconnect via Silicon Interposer หรือ Chiplet Architecture

---

6. ประโยชน์ทางอุตสาหกรรม (Industrial Applicability)

1. **Internet of Things (IoT)**: อุปกรณ์ Edge ที่ใช้แบตเตอรี่ขนาดเล็ก สามารถทำงาน AI inference ต่อเนื่องได้นานกว่า 1 ปี โดยไม่ต้องชาร์จ
2. **หุ่นยนต์และระบบอัตโนมัติ**: ประมวลผลภาพ เสียง และการเคลื่อนไหวแบบ Real-time ด้วยพลังงานต่ำ เหมาะกับหุ่นยนต์เคลื่อนที่
3. **การแพทย์ (Medical Devices)**: อุปกรณ์ฝังในร่างกาย (Implantable) เช่น เครื่องตรวจวัดคลื่นสมอง (EEG) แบบพกพา หรือ ประสาทเทียม (Neural Prosthetics)
4. **ยานยนต์ (Automotive)**: ระบบขับเคลื่อนอัตโนมัติ (ADAS) ที่ต้องการความหน่วงต่ำ (Low Latency) และความปลอดภัยสูง

5. **การป้องกันประเทศ (Defense)**: ระบบตรวจจับและพยากรณ์เป้าหมายแบบ Real-time ที่ทนทานต่อการโจมตีทางไฟฟ้าและแม่เหล็กไฟฟ้า
6. **Data Center**: ลดค่าใช้จ่ายด้านพลังงานของ AI Inference Server ลงได้สูงสุดถึง 100 เท่า เมื่อเทียบกับ GPU

---

7. ข้อถือสิทธิ (Patent Claims)

ข้อถือสิทธิที่ 1 (Independent Claim — Hardware Architecture):

สถาปัตยกรรมหน่วยประมวลผลนิวโรมอร์ฟิก ประกอบด้วย:

(ก) ชั้นฐาน CMOS (CMOS Base Layer) ที่ประกอบด้วยวงจร Mixed-Signal CMOS สำหรับทำหน้าที่เป็นเซลล์ประสาทเทียมแบบ Predictive Integrate-and-Fire (PIF) จำนวนหลายเซลล์เรียงตัวกันเป็นอาร์เรย์ โดยแต่ละเซลล์ประสาทดังกล่าวมีวงจรคำนวณอนุพันธ์อันดับหนึ่ง (dV/dt) และอนุพันธ์อันดับสอง (d^2V/dt^2) ของศักย์เยื่อหุ้มเซลล์ (Membrane Potential) เพื่อใช้ในการพยากรณ์ค่าศักย์เยื่อหุ้มเซลล์ในอนาคต

(ข) ชั้นหน่วยความจำแบบต้านทานแปรค่า (Memristive Layer) อย่างน้อยหนึ่งชั้น วางซ้อนอยู่ด้านบนของชั้นฐาน CMOS ด้วยเทคนิค Back-End-Of-Line (BEOL) ประกอบด้วยอาร์เรย์ไขว้สามมิติ (3D Crossbar Array) ของเซลล์หน่วยความจำแบบต้านทานแปรค่า (Memristor Cell) ที่ทำหน้าที่เป็นจุดเชื่อมต่อประสาทเทียม (Hardware Synapse) ที่เก็บค่าน้ำหนักแบบไม่ลบเลือน (Non-volatile Weight)

(ค) เครือข่ายส่งสัญญาณแบบไม่พึ่งสัญญาณนาฬิกา (Asynchronous Routing Network) ที่ใช้การแทนที่เหตุการณ์ด้วยแอดเดรส (Address Event Representation — AER) เพื่อส่งสัญญาณ Spike ระหว่างเซลล์ประสาท โดยส่วนที่ไม่มีการส่งสัญญาณจะถูกตัดกระแสไฟฟ้า (Power Gating) ให้ใช้พลังงานเป็นศูนย์

โดยที่สถาปัตยกรรมดังกล่าวไม่ต้องเข้าถึงหน่วยความจำภายนอกชิป (Zero Off-chip Memory Access) สำหรับการดำเนินการคำนวณของโครงข่ายประสาทเทียม

ข้อถือสิทธิที่ 2 (Dependent Claim — PIF Mechanism):

สถาปัตยกรรมตามข้อถือสิทธิที่ 1 ซึ่งเซลล์ประสาทเทียมแบบ PIF มีกลไกการยิงสัญญาณล่วงหน้า (Pre-emptive Firing) โดยวงจรพยากรณ์ (Predictive Circuit) จะคำนวณค่า $V_pred(t + \Delta t_pred)$ ตาม

สมการ:

```
V_pred = V_mem + α·dV/dt + β·d²V/dt² + γ·∫I_in·dt
```

และจะยิงสัญญาณ Spike ก่อนที่ศักย์เยื่อหุ้มเซลล์จริง (V_mem) จะถึงค่า Threshold (V_threshold) หาก V_pred มีค่าเท่ากับหรือมากกว่า V_threshold

ข้อถือสิทธิที่ 3 (Dependent Claim — Adaptive Prediction Window):

สถาปัตยกรรมตามข้อถือสิทธิที่ 2 ซึ่งค่า Δt_pred (Prediction Window) เป็นค่าที่ปรับเปลี่ยนได้ตามความสัมพันธ์:

```
Δt_pred = (V_threshold - V_mem) / (dV/dt + ε)
```

โดยที่ ε คือค่าคงที่เพื่อป้องกันการหารด้วยศูนย์

ข้อถือสิทธิที่ 4 (Dependent Claim — Dual Firing Mode):

สถาปัตยกรรมตามข้อถือสิทธิที่ 1 ซึ่งเซลล์ประสาทแต่ละเซลล์มีสองโหมดการทำงาน (Dual Firing Mode) ได้แก่:

(ก) โหมดพยากรณ์ (Predictive Mode): เซลล์ประสาทยิงสัญญาณล่วงหน้า (Pre-emptive Spike) เมื่อค่า V_pred ถึงค่า Threshold

(ข) โหมดตอบสนอง (Reactive Mode): เซลล์ประสาทยิงสัญญาณปกติ (Normal Spike) เมื่อค่า V_mem ถึงค่า Threshold ในกรณีที่ Predictive Mode ไม่ทำงาน

โดยทั้งสองโหมดสามารถสลับการทำงานได้แบบ Real-time โดยไม่ต้องหยุดการประมวลผล

ข้อถือสิทธิที่ 5 (Dependent Claim — 3D Memristor Crossbar):

สถาปัตยกรรมตามข้อถือสิทธิที่ 1 ซึ่งชั้นหน่วยความจำแบบต้านทานแปรค่าประกอบด้วยชั้น Crossbar Array จำนวน 4-16 ชั้นซ้อนกัน (Multi-layer Stacking) โดยแต่ละชั้นประกอบด้วย Wordline และ Bitline ในแนวตั้งฉากกัน โดยมีเซลล์ Memristor แบบ 1T1R (One Transistor One Resistor) อยู่ที่จุดตัด (Crosspoint) และมีชั้นฉนวนระหว่างชั้น (Inter-Layer Dielectric) คั่นระหว่างชั้น

ข้อถือสิทธิที่ 6 (Dependent Claim — STDP Online Learning):

สถาปัตยกรรมตามข้อถือสิทธิที่ 1 ซึ่งมีวงจรเรียนรู้แบบ Spike-Timing-Dependent Plasticity (STDP) ในระดับฮาร์ดแวร์ที่ปรับค่าน้ำหนักของ Memristor แบบ Real-time โดยการตรวจจับเวลาต่าง (Δt) ระหว่าง Pre-synaptic Spike และ Post-synaptic Spike และจ่ายแรงดัน Set/Reset Pulse ที่เหมาะสมให้กับ Memristor Cell

ข้อถือสิทธิที่ 7 (Dependent Claim — AER Routing with Power Gating):

สถาปัตยกรรมตามข้อถือสิทธิที่ 1 ซึ่งเครือข่าย AER Routing ประกอบด้วย:

(ก) คำขอเส้นทาง (Path Request): เมื่อมี Spike ถูกสร้างขึ้น AER Encoder จะสร้างแพ็กเก็ตแอดเดรส (Address Packet) ที่มี Source ID และ Timestamp

(ข) การจัดการเส้นทาง (Routing Arbitration): และส่งแพ็กเก็ตดังกล่าวผ่าน Router Node ซึ่งมี Arbiter สำหรับจัดการ Collision

(ค) การตัดกระแสส่วนที่ไม่ใช้ (Power Gating): โดย Router Node ที่ไม่มีการใช้งานในขณะนั้นจะถูกตัดกระแส (Idle State) ด้วย Power Gating Switch และจะกลับมาจ่ายกระแส (Active State) ภายในเวลาไม่เกิน 1 นาโนวินาทีเมื่อมี Spike Packet มาถึง

ข้อถือสิทธิที่ 8 (Independent Claim — Method Claim):

วิธีการประมวลผลสัญญาณประสาทเทียมแบบ Predictive Integrate-and-Fire (PIF) ประกอบด้วยขั้นตอน:

(ก) รับกระแสอินพุต (Input Current) จากอาร์เรย์หน่วยความจำแบบต้านทานแปรค่าที่จุดเชื่อมต่อประสาทเทียม (Hardware Synapse) เข้าสู่วงจรรวมกระแส (Current Integrator) เพื่อสะสมเป็นศักย์เยื่อหุ้มเซลล์ (Membrane Potential, V_mem)

(ข) คำนวณอัตราการเปลี่ยนแปลงของศักย์เยื่อหุ้มเซลล์ (dV/dt) ด้วยวงจรอนุพันธ์อันดับหนึ่ง (First Derivative Circuit) และคำนวณความเร่งของศักย์เยื่อหุ้มเซลล์ (d^2V/dt^2) ด้วยวงจรอนุพันธ์อันดับสอง (Second Derivative Circuit)

(ค) คำนวณค่าพยากรณ์ศักย์เยื่อหุ้มเซลล์ในอนาคต (V_pred) โดยใช้ค่าดังกล่าวร่วมกับสัมประสิทธิ์ที่กำหนดไว้ (α, β, γ)

(ง) เปรียบเทียบค่า V_pred กับค่า Threshold (V_threshold) โดยใช้วงจรเปรียบเทียบ (Comparator)

(จ) หาก V_pred มีค่าเท่ากับหรือมากกว่า V_threshold ให้สร้างสัญญาณ Spike ล่วงหน้า (Pre-emptive Spike) และรีเซ็ตศักย์เยื่อหุ้มเซลล์

(ฉ) หาก V_pred มีค่าน้อยกว่า V_threshold แต่ V_mem มีค่าเท่ากับหรือมากกว่า V_threshold ให้สร้างสัญญาณ Spike ปกติ (Normal Spike) และรีเซ็ตศักย์เยื่อหุ้มเซลล์

ข้อถือสิทธิที่ 9 (Dependent Method Claim — Decoupled Prediction):

วิธีการตามข้อถือสิทธิที่ 8 ซึ่งขั้นตอน (ค)-(จ) ทำงานแบบคู่ขนานและอิสระ (Decoupled) จากการรวมประจุในขั้นตอน (ก) ทำให้การพยากรณ์ไม่รบกวนการสะสมประจุของ Membrane Potential โดยใช้ Sample-and-Hold Circuit ในการ sampling ค่า V_mem, dV/dt, d^2V/dt^2

ข้อถือสิทธิที่ 10 (Independent Claim — Security Hardware):

ระบบรักษาความปลอดภัยระดับฮาร์ดแวร์สำหรับสถาปัตยกรรมหน่วยประมวลผลนิวโรมอร์ฟิก ประกอบด้วย:

(ก) วงจรสร้างสัญญาณ Spike ปลอม (Decoy Spike Generator) ที่สร้างสัญญาณ Spike ที่มีลักษณะทางไฟฟ้าเหมือนสัญญาณ Spike จริงแต่ไม่มีข้อมูล กระจายตัวทั่วทั้งชิป

(ข) วงจรสุ่มเส้นทางเดินสัญญาณ (Obfuscated Routing Path) ที่ใช้ Physical Unclonable Function (PUF) ในการกำหนดเส้นทางการเดินของสัญญาณจริง ทำให้ยากต่อการวิเคราะห์ Reverse Engineering

(ค) วงจรสุ่มสัญญาณรบกวนแหล่งจ่ายไฟ (Power Profile Randomizer) ที่สร้างสัญญาณรบกวนกระแสสลับ (AC Noise) เพื่อบดบังลายเซ็นการใช้พลังงาน (Power Signature) ของการคำนวณจริง ป้องกันการโจมตีแบบ Differential Power Analysis (DPA)

(ง) ชั้นตาข่ายป้องกันการแกะชิป (Anti-Tamper Mesh) ที่ไวต่อการขาด (Continuity Sensing) โดยเมื่อมีผู้พยายามเจาะหรือกัดกร่อนชิป ตาข่ายจะขาดและสั่งให้วงจรรีเซ็ตลบข้อมูลในหน่วยความจำ Memristor Array ทั้งหมด

ข้อถือสิทธิที่ 11 (Dependent Claim — AER Security):

สถาปัตยกรรมตามข้อถือสิทธิที่ 1 ซึ่งแพ็กเก็ต AER Address ประกอบด้วยฟิลด์ Priority Flag (2 บิต) สำหรับระบุระดับความสำคัญของ Spike โดย Spike ที่มี Priority สูงจะได้รับการจัดสรรเส้นทางก่อน (Preemptive Routing) ซึ่งรวมถึง Spike ปลอม (Decoy Spike) ที่ถูกกำหนด Priority ต่ำสุดเพื่อไม่ให้รบกวนการทำงานของ Spike จริง

ข้อถือสิทธิที่ 12 (Dependent Claim — Fabrication):

วิธีการผลิตสถาปัตยกรรมหน่วยประมวลผลนิวโรมอร์ฟิกตามข้อถือสิทธิที่ 1 ประกอบด้วยขั้นตอน:

(ก) สร้างชั้นฐาน CMOS ด้วยกระบวนการ CMOS มาตรฐานที่เทคโนโลยีระดับ 28nm หรือเล็กกว่า บน Silicon Wafer โดยบรรจุวงจร Mixed-Signal CMOS สำหรับ PIF Neuron, AER Router, STDP Circuit, และ Security Circuit

(ข) สร้างชั้นหน่วยความจำแบบต้านทานแปรค่าอย่างน้อยหนึ่งชั้นด้านบนของชั้นฐาน CMOS ด้วยเทคนิค Back-End-Of-Line (BEOL) โดยเริ่มจากการสะสม Bottom Electrode ด้วย TiN/TaN, ชั้น Switching ด้วย Metal-Oxide (HfO_2, TaO_x), และ Top Electrode ด้วย TiN/Pt หรือ TiN/Ti

(ค) สร้าง Inter-Layer Dielectric (ILD) เพื่อเป็นฉนวนระหว่างชั้น Crossbar

(ง) ทำซ้ำขั้นตอน (ข)-(ค) สำหรับ Crossbar Layer ถัดไปตามจำนวนชั้นที่ต้องการ

(จ) สร้างชั้นระบบระบายความร้อนและป้องกันคลื่นแม่เหล็กไฟฟ้า (Micro-cooling & EM Shielding Layer) และ Anti-Tamper Mesh ชั้นบนสุด

ข้อถือสิทธิที่ 13 (Dependent Claim — Non-volatile Weight Storage):

สถาปัตยกรรมตามข้อถือสิทธิที่ 1 ซึ่งค่าน้ำหนัก (Weights) ที่เก็บใน Memristor Crossbar Array ยังคงอยู่แม้ไม่มีพลังงานจ่ายให้กับชิป (Non-volatile) ทำให้ไม่ต้อง Boot Load โมเดลเมื่อเริ่มต้นทำงาน (Instant-on)

ข้อถือสิทธิที่ 14 (Dependent Claim — Multi-energy Domain):

สถาปัตยกรรมตามข้อถือสิทธิที่ 1 ซึ่งสามารถทำงานในโหมดประหยัดพลังงาน (Low Power Mode) โดยการลดจำนวนชั้น Memristor Stack ที่ทำงานพร้อมกัน หรือลดความแม่นยำของ PIF Prediction (ลด Precision ของ α, β, γ) เพื่อยืดอายุแบตเตอรี่

ข้อถือสิทธิที่ 15 (Independent Claim — สิทธิบัตรกระบวนการ):

กระบวนการประมวลผลสัญญาณในสถาปัตยกรรมนิวโรมอร์ฟิกแบบหลายเวลา (Multi-temporal Processing Method) ประกอบด้วยขั้นตอน:

(ก) แยกสัญญาณอินพุตขาเข้าออกเป็นเฟรมย่อยตามเวลา (Temporal Frames) โดยใช้ Shift Register แบบ Asynchronous

(ข) ประมวลผลเฟรมย่อยแต่ละเฟรมขนานกันใน PIF Neuron แต่ละตัวที่เชื่อมต่อถึงกันด้วย AER Router

(ค) ใช้กลไก PIF รวมผลการพยากรณ์จากทุกเฟรมในลักษณะ Time-warped Integration โดย Spike ที่พยากรณ์จากเฟรมอนาคต (Future Frame) จะมีผลต่อการ integrate ของเฟรมปัจจุบัน (Present Frame) แบบทันทีทันใด

(ง) สร้าง Output Spike ที่ encode ข้อมูลจากหลายช่วงเวลาพร้อมกันใน Spike เดียว (Temporal Encoding)

---

8. บทสรุปและข้อแนะนำทางยุทธศาสตร์ (Strategic Summary)

8.1 สาระสำคัญที่สามารถจดสิทธิบัตรได้

ข้อ	องค์ประกอบที่จดสิทธิบัตรได้	ความใหม่ (Novelty)	ลักษณะทางเทคนิค
1	Predictive Integrate-and-Fire (PIF)	★★★	Derivative-based prediction
2	3D Memristor + CMOS Monolithic Integration	★★★	BEOL stacking

ข้อ	องค์ประกอบที่จดสิทธิบัตรได้	ความใหม่ (Novelty)	ลักษณะทางเทคนิค
3	Asynchronous AER + Power Gating	★★	Event-driven
4	Decoy Routing + PUF Obfuscation	★★★	Hardware security
5	Dual-Mode Firing (Predictive + Reactive)	★★★	Adaptive firing
6	Anti-Tamper Mesh with Memristor Erase	★★★	Self-destruct memory

8.2 คำแนะนำก่อนยื่น

1. **ค้นหา Prior Art (Patent Search)**: ควรทำการค้นหาสิทธิบัตรที่เกี่ยวข้องโดยเฉพาะจาก USPTO, JPO, CNIPA และ WIPO ก่อนยื่น
2. **เขียน Claims แบบหลายชั้น**: แยกชุด Claims สำหรับยื่นในประเทศไทย (Thai Patent) และต่างประเทศ (PCT/USPTO) โดยเฉพาะ Claims เกี่ยวกับกลไก PIF ควรใช้ภาษาให้กว้างที่สุด (Broadest Reasonable Interpretation)
3. **เขียน Provisional Application (US) ก่อน**: หากต้องการยื่นในสหรัฐฯ ควรยื่น Provisional Patent Application ก่อนเพื่อให้ได้ Filing Date ก่อน แล้วค่อยยื่น PCT ภายใน 12 เดือน
4. **เตรียม Diagram/Figure**: สำนักงานสิทธิบัตรต้องการ Diagram ของ:
    - Figure 1: Block Diagram ของสถาปัตยกรรมโดยรวม
    - Figure 2: Schematic ของ PIF Neuron Circuit
    - Figure 3: Timing Diagram แสดง Pre-emptive Spike
    - Figure 4: Cross-section View แสดง Layer Stack
    - Figure 5: 3D Memristor Crossbar Array Structure
    - Figure 6: AER Router และ Data Flow

- Figure 7: Decoy Routing และ Security Circuit

5. **เตรียม Simulation Results**: ควรมีผลการจำลอง (Simulation) เช่น SPICE Simulation หรือ Benchmark ผลเทียบกับ LIF แบบดั้งเดิมเพื่อแสดงให้เห็นถึงประโยชน์ (Utility) ของการประดิษฐ์

6. **พิจารณายื่นประเภทสิทธิบัตร**: เนื่องจากมีการประดิษฐ์หลายอย่างร่วมกัน ควรพิจารณายื่นหลายฉบับ (Patent Family) เช่น:

   - ฉบับที่ 1: สถาปัตยกรรมฮาร์ดแวร์ (Device/Hardware Patent)
   - ฉบับที่ 2: วิธีการประมวลผล PIF (Method Patent)
   - ฉบับที่ 3: ระบบความปลอดภัยฮาร์ดแวร์ (Security System Patent)
   - ฉบับที่ 4: กระบวนการผลิต (Fabrication Method Patent)

เอกสารแนบ (ที่จะต้องเตรียมเพิ่มเติม)

1. ☐ รูปเขียน (Drawings/Figures) - จำนวน 7 รูป
2. ☐ ผลการทดลอง/Simulation
3. ☐ บทสรุปข้อถือสิทธิ (Abstract)
4. ☐ คำรับรองผู้ประดิษฐ์ (Inventor Declaration)
5. ☐ หนังสือรับรองการโอนสิทธิ (หากจดโดยนิติบุคคล)

จัดทำโดย: ไชยภพ นิลแพทย์ (The Architect) วันที่: [วันที่ยื่น]

เอกสารนี้เป็นร่างสำหรับปรึกษาทนายสิทธิบัตรก่อนยื่นจริง ข้อมูลทางเทคนิคอาจต้องปรับปรุงตามคำแนะนำของผู้เชี่ยวชาญด้านสิทธิบัตร